A method for the early health technology assessment of novel biomarker measurement in primary prevention programs

Postmus, Douwe; de Graaf, Gimon; Hillege, Hans L.; Steyerberg, Ewout W.; Buskens, Erik

Published in:
Statistics in Medicine

DOI:
10.1002/sim.5434

Publication date:
2012

Research Article

Statistics in Medicine

Received 21 February 2011, Accepted 12 April 2012 Published online 16 July 2012 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.5434

A method for the early health technology assessment of novel biomarker measurement in primary prevention programs

Douwe Postmus,[a*†] Gimon de Graaf,[a] Hans L. Hillege,[a,b] Ewout W. Steyerberg[c] and Erik Buskens[a]

Many promising biomarkers for stratifying individuals at risk of developing a chronic disease or subsequent complications have been identified. Research into the potential cost-effectiveness of applying these biomarkers in actual clinical settings has however been lacking. Investors and analysts may improve their venture decision making should they have indicative estimates of the potential costs and effects associated with a new biomarker technology already at the early stages of its development. To assist in obtaining such estimates, this paper presents a general method for the early health technology assessment of a novel biomarker technology. The setting considered is that of primary prevention programs where initial screening to select high-risk individuals eligible for a subsequent intervention occurs, for example, prevention of type 2 diabetes. The method is based on quantifying the health outcomes and downstream healthcare consumption of all individuals who get reclassified as a result of moving from a screening variant based on traditional risk factors to a screening variant based on traditional risk factors plus a novel biomarker. As these individuals form well-defined subpopulations, a combination of disease progression modeling and sensitivity analysis can be used to perform an initial assessment of the maximum increase in screening cost for which the use of the new biomarker technology is still likely to be cost effective.

Keywords: biomarker; cost-effectiveness analysis; early health technology assessment; primary prevention

1. Introduction

Much research effort is currently directed at discovering novel biomarkers for identifying individuals at risk of developing a chronic disease (primary prevention) or subsequent complications (tertiary prevention). As these biomarkers provide additional information beyond standard clinical risk factors, applying them in actual clinical settings is expected to result in improved risk stratification. This, in turn, may help to optimize the selection of individuals eligible for a focused intervention, such as behavioral counseling or chemoprevention. Ultimately, this should improve the population's health outcomes at affordable (possibly lower) costs.

After a promising biomarker has been identified and a (prototype) technology has been developed to measure this biomarker in actual clinical settings, its performance needs to be critically evaluated before the new biomarker technology will eventually be adopted in clinical practice. According to Hlatky *et al.* [1], such a critical assessment involves six phases, ranging from showing that the levels of the novel biomarker differ between individuals with and without the outcome of interest (proof of concept) to assessing whether using the biomarker improves health outcomes at an affordable cost

[a] *Department of Epidemiology, University Medical Center Groningen, University of Groningen, The Netherlands*
[b] *Department of Cardiology, University Medical Center Groningen, University of Groningen, The Netherlands*
[c] *Department of Public Health, Erasmus Medical Center, Rotterdam, The Netherlands*
**Correspondence to: Douwe Postmus, Department of Epidemiology, University of Groningen, University Medical Center Groningen, PO Box 30001, 9700 RB Groningen, The Netherlands.*
[†] *E-mail: d.postmus@umcg.nl*

(cost-effectiveness). In this traditional framework, cost-effectiveness analysis is conducted at the end of the product development process. The results are intended to assist health policy makers in deciding whether the new biomarker technology should be routinely adopted in clinical practice. This form of health technology assessment (HTA) is what Pietzsch and Paté-Cornell [2] referred to as classical HTA, to be distinguished from an initial assessment of the likely costs and effects associated with a new medical technology at the early stages of its development process. This so-called early HTA is conducted before the technology has been fully developed and serves to support health technology firms in making appropriate product investment decisions.

Although appropriate quantification of the added predicted value of a novel biomarker over conventional risk factors is a problem of active research and debate [3–9], research into the cost-effectiveness of applying such a biomarker in actual clinical settings has so far been limited to two recent case studies in the context of prioritizing patients waiting for coronary artery surgery [10, 11]. A more in-depth discussion on how the cost-effectiveness of using prognostic biomarkers could be established seems therefore desirable. To that end, this paper presents a general method for the early HTA of a novel biomarker technology that is used, in combination with a set of conventional risk factors, as an initial screening test to select high-risk individuals eligible for a subsequent preventive intervention. We illustrate the use of the method in a case study related to the prevention of type 2 diabetes.

2. Added predictive value and cost-effectiveness of novel biomarker measurement in primary prevention programs

IJzerman and Steuten [12] have recently provided a conceptual model of the medical technology development process. According to their model, this process consists of four main stages: (i) basic research, (ii) translational research, (iii) clinical research, and (iv) market access. Preceding each of these stages is a decision gate where it has to be decided whether to proceed with the next stage, and if so, in what direction. In this paper, we assume that the basic research on biological mechanisms is completed and that this has resulted in the identification of several candidate biomarkers. We are therefore at the decision gate preceding the translational research phase, where it has to be decided which of these biomarkers should be selected for further development, if any. The purpose of performing early HTA at this stage of the product development process is to assist health technology firms in making realistic commercial valuations of the conceived new products by providing for each potential new biomarker technology a rough estimate of the maximum additional cost for which its intended clinical application is still likely to be cost effective. In this section, we will describe how this upper bound on the technology's cost, also known as the commercial headroom available [13], can be estimated for an improved, biomarker-based screening test.

Consider N individuals who participate in a primary prevention program. On the basis of the results of an initial screening, individuals are classified into m ordinal risk categories, such as low, intermediate, and high risk in case of three categories. Those who are considered to be at risk are offered a subsequent intervention, which may be tailored to the risk category an individual is classified into (e.g., no intervention in individuals who are being classified as low risk, a non-invasive and relatively safe intervention in individuals who are being classified as intermediate risk, and an invasive and more risky intervention in individuals who are being classified as high risk). Suppose that a decision maker can choose between two variants of the risk stratification model: screening variant s^1 in which the risk stratification is based on a vector of cutoff points $\gamma^1 = \left(\gamma_1^1, \ldots, \gamma_{m-1}^1\right)'$ on a risk score consisting of conventional risk factors and screening variant s^2 in which the risk stratification is based on a vector of cutoff points $\gamma^2 = \left(\gamma_1^2, \ldots, \gamma_{m-1}^2\right)'$ on a risk score comprising the same conventional risk factors as well as a novel biomarker. For clinically meaningful values of the cutoff points γ^1 and γ^2, consider the m by m reclassification table that results from combining the risk classifications obtained under s^1 and s^2 (Table I). Let N_{kl} denote the number of individuals within the klth entry of the reclassification table (i.e., all individuals who become classified in the kth risk category under s^1 and in the lth risk category under s^2). It then follows that a fraction of $\frac{\sum_{k=1}^{m} \sum_{l \neq k} N_{kl}}{N}$ of the individuals are reclassified when applying s^2 instead of s^1.

The extent to which this reclassification can be considered an improvement can be determined in several ways [14]. A measure of reclassification that has nowadays gained wide-spread acceptance is the net reclassification index (NRI) [7]. The main idea behind this measure is to consider the reclassification

Table I. The reclassification table that results from combining the risk classifications under s^1 and s^2.

		Classification under s^2			
		Risk category 1	Risk category 2	...	Risk category m
Classification under s^1	Risk category 1	N_{11}	N_{12}	...	N_{1m}
	Risk category 2	N_{21}	N_{22}	...	N_{2m}
	⋮	⋮	⋮	⋱	⋮
	Risk category m	N_{m1}	N_{m2}	...	N_{mm}

of individuals who develop and who do not develop the event of interest separately. Moving from s^1 to s^2 can then be considered an improvement when the proportion of subjects who move upward toward a higher risk category is larger than the proportion of subjects who move downward toward a lower risk category for individuals with the event and when the opposite holds for individuals without the event. To assess this in a formal way, consider a random sample of size n from the screening population, and let n_{kl}^{event} and $n_{kl}^{\text{no event}}$ denote the number of events and non-events within the klth cell of the reclassification table corresponding to this sample, respectively. The NRI is then computed as

$$\frac{\sum_{k=1}^{m}\sum_{l>k} n_{kl}^{\text{event}} - \sum_{k=1}^{m}\sum_{l<k} n_{kl}^{\text{event}}}{n^{\text{event}}} + \frac{\sum_{k=1}^{m}\sum_{l<k} n_{kl}^{\text{no event}} - \sum_{k=1}^{m}\sum_{l>k} n_{kl}^{\text{no event}}}{n^{\text{no event}}},$$

where n^{event} and $n^{\text{no event}}$ denote the total number of events and non-events in the total sample, respectively. The novel biomarker is then considered to have incremental predictive value over the conventional risk factors if the null hypothesis of NRI $= 0$ is rejected.

Although the NRI and other proposed measures of reclassification are useful for establishing the added predicted ability of a novel biomarker, they do not directly provide insight into which of the risk stratification models would be preferable from a societal perspective. To address this latter aspect, we have to determine whether the increase in added predictive value is sufficient to make changing from s^1 to s^2 an efficient allocation of scarce healthcare resources, and this is the domain of health economic analysis [15]. In this type of analysis, it is common practice to assume that all relevant health effects can be aggregated into a single measure of effectiveness. The net monetary benefit (NMB) of an intervention can then be calculated by assuming a threshold value of the decision maker's willingness to pay for one unit of health gain [15, 16]. The most common measure of effectiveness is the quality-adjusted life year (QALY), and this is also the one that will be used in the case study. In the remainder of this paper, we will therefore assume that the effectiveness of the two screening variants is evaluated in terms of QALYs. The equations derived in this section are however applicable in all situations where the health consequences are captured in terms of a single measure of effectiveness.

To determine the relative merits of the two screening variants, let c^i and e^i denote the average cost and QALYs associated with screening variant s^i, and let λ denote the willingness to pay (in terms of monetary units) for a QALY. Screening variant s^2 is then preferred over screening variant s^1 if

$$\text{NMB}^2 - \text{NMB}^1 \geqslant 0, \tag{1}$$

where $\text{NMB}^i = \lambda e^i - c^i$ represents the average NMB associated with screening variant s^i. To use Equation (1) to compute the headroom available to the improved, biomarker-based screening test, let the treatment assignment indicator t_{kl}^i return the treatment that is assigned to the individuals in the klth entry of the reclassification table under screening variant s^i. For example, if under s^1 treatment A is offered to all individuals who are classified into risk category 2, $t_{2l}^1 =$ treatment A, $\forall l \in \{1, \ldots, m\}$. The average cost and QALYs associated with screening variant s^i can then be written as

$$c^i = c_{\text{scr}}^i + \frac{1}{N}\sum_{k=1}^{m}\sum_{l=1}^{m} N_{kl} c_{kl}^{t_{kl}^i}, \tag{2}$$

$$e^i = \frac{1}{N}\sum_{k=1}^{m}\sum_{l=1}^{m} N_{kl} e_{kl}^{t_{kl}^i}, \tag{3}$$

where $c_{kl}^{t_{kl}^i}$ and $e_{kl}^{t_{kl}^i}$ denote the average cost and QALYs associated with applying treatment t_{kl}^i to the individuals in the klth entry of the reclassification table, and where c_{scr}^i denotes the average screening cost under screening variant s^i. If we now define $f_{kl} = N_{kl}/N$ and $\Delta_{\text{scr}} = c_{\text{scr}}^2 - c_{\text{scr}}^1$, it follows by substituting (2) and (3) into (1) that screening variant s^2 is preferred over screening variant s^1 if

$$\Delta_{\text{scr}} \leqslant \sum_{k=1}^{m} \sum_{j \neq k} f_{kl} \left[\lambda \left(e_{kl}^{t_{kl}^2} - e_{kl}^{t_{kl}^1} \right) - \left(c_{kl}^{t_{kl}^2} - c_{kl}^{t_{kl}^1} \right) \right]. \tag{4}$$

Individuals who end up at one of the diagonal entries of the reclassification table are assigned to the same treatment under both s^1 and s^2. Consequently, it is not required to consider these individuals' QALYs and downstream healthcare consumption when choosing between the two screening variants. In Equation (4), this is reflected by the fact that the average beneficial and/or harmful consequences (in terms of downstream NMB) associated with switching from s^1 to s^2, that is, $\lambda \left(e_{kl}^{t_{kl}^2} - e_{kl}^{t_{kl}^1} \right) - \left(c_{kl}^{t_{kl}^2} - c_{kl}^{t_{kl}^1} \right)$, are only computed for those individuals who move upwards or downwards in risk classification. For the biomarker-based screening variant to be cost effective, the overall increase in downstream NMB (i.e., the right-hand side of Equation (4)) needs to be sufficiently large to offset the upfront increase in screening cost, which are incurred by all individuals, irrespective of whether they are reclassified.

3. Parameter estimation

In this section, we will describe how the parameters at the right-hand side of Equation (4) can be estimated at the decision gate preceding the translational research stage of the medical technology development process. As the amount of clinical data available for estimating these parameters very much depends on whether the biomarker in question has already been measured in a prospective cohort study, we will make a distinction between technologies that aim at providing an alternative way of measuring an existing biomarker and technologies that aim at measuring a completely new biomarker.

3.1. General considerations

As the initiation of a preventive intervention is expected to have cost and effect implications on the remainder of a patient's live, the appropriate time horizon for the economic evaluation of such interventions is the patient's lifetime [15]. In such situations, health economic analysts generally rely on disease progression modeling to extrapolate from the event rates and treatment effects observed in clinical trials and observational studies to what would be expected to happen over a lifetime [16]. The quantitative models used for this purpose typically consist of several discrete health states reflecting the occurrence of the events of interest and a set of transition intensities (or transition probabilities in case of a discrete-time model) that govern the movement between these health states. The expected long-term cost and effect consequences of an intervention can then be estimated by multiplying the average sojourn time in each of the model's health states by a cost and utility weight attached to these health states. To include patient heterogeneity into the model, the logarithms of the transition intensities are sometimes expressed as linear functions of a set of explanatory covariates, resulting in a so-called patient-level model. Disease progression models that do not take into account patient heterogeneity are generally referred to as cohort models [17].

For the individuals in the klth cell of the reclassification table, the expected cost and QALY consequences of moving from s^1 to s^2 will depend on three main aspects: (i) the cumulative disease incidence $I_{kl}(\tau)$ as a function of the time since screening τ that would be observed in this population in the absence of screening; (ii) the reduction in cumulative disease incidence due to t_{kl}^1; and (iii) the relative risk of t_{kl}^2 relative to t_{kl}^1. Our strategy is to derive the health economic consequences that result from these changes in the cumulative disease incidence through disease progression modeling. As the individuals from the different cells of the reclassification table form well-defined subpopulations, we propose to fit separate disease progression models to each of these subpopulations. In particular, let the vector α_{kl} denote the model parameters that apply to subpopulation kl. The disease progression models can then be expressed as functions $h\left(t_{kl}^i | \alpha_{kl}\right)$ that map the administered treatment to the expected values of $c_{kl}^{t_{kl}^i}$ and $e_{kl}^{t_{kl}^i}$.

3.2. Prospective data available

Many companies in the medical device industry do not only focus on developing novel equipment for measuring promising new biomarkers but also on finding alternative (e.g., more efficient, less invasive, or less risky) ways of measuring an existing biomarker, such as a multiplex ELISA that can simultaneously measure a whole panel of biomarkers. In such situations, it may already be possible to evaluate the added predictive value of the selected (panel of) biomarker(s) over a set of conventional risk factors by applying the currently available measurement techniques to blood, urine, or tissue samples collected in an existing cohort study. Base-case values of the fractions f_{kl} can then readily be derived from a reclassification table that is constructed from the data collected in this study. The same applies to all parameters in α_{kl} that directly depend on the incremental predicted value of the considered biomarker(s).

3.3. Prospective data not available

When dealing with a completely new biomarker, nothing will yet be known about the performance of this biomarker in actual clinical settings. Initial values of f_{kl} must then be derived from surrogate data, such as early bench and animal testing, the performance of related but already clinically validated biomarkers, or expert judgment. A similar problem is encountered when specifying the parameters of the disease progression model: although it may be possible to obtain some of the parameter values from previously conducted economic evaluations, such as the costs and utilities attached to the model's transient health states, others depend on the incremental predictive value of the new biomarker and must therefore also be derived from indirect sources. Probability Aggregation for Medical Device Assessment [2] is particularly suited for synthesizing evidence from multiple indirect sources, such as the results from several pilot studies in different types of animal model. For a thorough discussion on how expert knowledge can be elicited and incorporated in a probabilistic way, we refer the reader to [18].

4. Initial economic evaluation

After the base-case values of all parameters have been specified, the commercial headroom available to the new biomarker technology, denoted by $\Delta_{\text{scr}}^{\max}$, can be estimated by applying the algorithm depicted in Figure 1. As the values of most parameters are still uncertain at the early stages of the medical technology development process, the base-case analysis should be followed by an extensive amount of sensitivity analysis to determine the robustness of the obtained results with respect to changes in the parameter values. How the sensitivity analysis can best be conducted depends on the amount of clinical data available [19]. If the added predictive value of the considered biomarker has already been evaluated in a prospective cohort study, the use of probabilistic sensitivity analysis (PSA) seems most appropriate as probability distributions of the parameters of interest can then directly be derived from the data collected within this study. On the other hand, unless expert knowledge has been elicited in a probabilistic way, the use of PSA is generally not feasible when the novel biomarker has not yet been measured in a prospective cohort study. The use of a deterministic approach, such as one-way sensitivity analysis, would then be more appropriate.

Input: $\lambda, f_{kl}, \alpha_{kl}, t_{kl}^1, t_{kl}^2, \quad k,l = 1,\ldots,m,\ k \neq l$
Output: $\Delta_{\text{scr}}^{\max}$
1: **for** $k \leftarrow 1$ **to** m **do**
2: **for** $l \leftarrow 1$ **to** m **do**
3: **if** $k \neq l$ **then**
4: $(c_{kl}^{t_{kl}^1}, e_{kl}^{t_{kl}^1}) \leftarrow h(t_{kl}^1 | \alpha_{kl})$
5: $(c_{kl}^{t_{kl}^2}, e_{kl}^{t_{kl}^2}) \leftarrow h(t_{kl}^2 | \alpha_{kl})$
6: **end if**
7: **end for**
8: **end for**
9: $\Delta_{\text{scr}}^{\max} \leftarrow \sum_{k=1}^{m} \sum_{j \neq k} f_{kl} \left[\lambda (e_{kl}^{t_{kl}^2} - e_{kl}^{t_{kl}^1}) - (c_{kl}^{t_{kl}^2} - c_{kl}^{t_{kl}^1}) \right]$

Figure 1. Algorithm for estimating $\Delta_{\text{scr}}^{\max}$ for a specific set of parameter values.

In a PSA, the uncertainty in one or more input parameters is propagated to uncertainty in the outcome variable by repeatedly calculating $\Delta_{\text{scr}}^{\max}$ for different samples from the (joint) distribution of the input parameters. As the parameters of interest are treated as random variables, it seems logical to adopt a Bayesian approach in estimating the probability distributions of these parameters. In particular, for given values of the cutoff points, consider a random sample $\psi_j = (y_j, \mathbf{z}_j)$, $j = 1, \ldots, n$, from the screening population, where y_j denotes the cell of the reclassification table where individual j is classified into and $\mathbf{z}_j$ all other measurements taken on individual j required to derive the joint distribution of the vector $\alpha = (\alpha_{11}, \ldots, \alpha_{mm})'$. The ψ_js may be assumed to be independent across individuals, but y_j and $\mathbf{z}_j$ are likely to be correlated within individuals. Let $g(\psi_j | f_{11}, \ldots, f_{mm}, \alpha)$ denote the joint density of ψ_j and consider the factorization

$$g(\psi_j | f_{11}, \ldots, f_{mm}, \alpha) = g(\mathbf{z}_j | y_j, \alpha) g(y_j | f_{11}, \ldots, f_{mm}) = g(\mathbf{z}_j | y_j, \alpha_{y_j}) g(y_j | f_{11}, \ldots, f_{mm}). \quad (5)$$

Marginally, $g(y_j | f_{11}, \ldots, f_{mm})$ is the frequency function of a discrete random variable with probability f_{kl} that individual j is classified into the klth cell of the reclassification table. Given that $y_j \perp\!\!\!\perp y_k$, the observed number of individuals $n_{11}, \ldots, n_{mm}$ in the different cells of the reclassification table is multinomially distributed with probability vector $(f_{11}, \ldots, f_{mm})'$. With a Dirch$(a_{11}, \ldots, a_{mm})$ conjugate prior used, the posterior distribution of $(f_{11}, \ldots, f_{mm})'$ can be modeled as $(f_{11}, \ldots, f_{mm})' \sim$ Dirch$(a_{11}+n_{11}, \ldots, a_{mm}+n_{mm})$ [20]. To estimate the joint distribution of $\alpha_{kl}, k, l = 1, \ldots, m, k \neq l$, it makes sense to condition the observations on the observed value of y_j. The data $\mathbf{z}_{kl,1}, \ldots, \mathbf{z}_{kl,n_{kl}}$ can then be treated as independent samples from each level of y_j, such that separate multivariate Bayesian models can be fitted to each of these samples to obtain the posterior distributions of α_{kl}.

In a deterministic sensitivity analysis, no attempt is made to specify parameter uncertainty through the use of probability distributions. Instead, reasonable lower and upper bounds are identified for each of the parameters of interest, after which the actual sensitivity analysis is conducted by exploring in a deterministic way how different combinations of the parameter values affect the value of $\Delta_{\text{scr}}^{\max}$. To explore which of the input parameters have the highest impact on the outcome variable, it is common practice to change one parameter value at a time, resulting in a so-called one-way sensitivity analysis. It is also possible to perform a multi-way sensitivity analysis by changing two or more parameter values simultaneously. However, the parameter values are still allowed to vary independently from each other as nothing is known about the correlation between these parameters.

5. Illustrative case study

To demonstrate how our proposed method can assist in quantifying the headroom available to an improved, biomarker-based screening test, we applied the method in a case study related to the prevention of type 2 diabetes mellitus (DM2) and its associated microvascular and macrovascular complications.

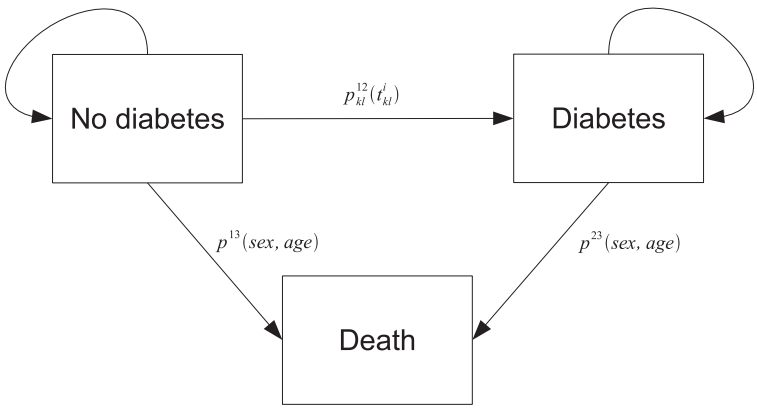

Figure 2. A discrete-time Markov model with three health states.

5.1. *Clinical context*

Lifestyle interventions have been previously shown to be cost-effective strategies to reduce the incidence of DM2 in patients with pre-diabetes [21]. The primary prevention program considered in this case study therefore consists of providing a lifestyle intervention to individuals who are at increased risk of

Table II. Model parameters, their initial values, and the sources used to obtain these values.

Parameter	Symbol in text (if defined)	Value	Source
Willingness to pay for a QALY	λ	20,000	Reference value
Reclassification table			
Size of the study population	n	4977	[22]
Size of subpopulation kl	n_{kl}	Table III	[22]
Observed number of DM2 cases in subpopulation kl	n_{kl}^{event}	Table III	[22]
Fraction of individuals in subpopulation kl	f_{kl}	n_{kl}/n	
Seven-year incidence of DM2 in subpopulation kl in the absence of screening	$I_{kl}(7)$	$n_{kl}^{\text{event}}/n_{kl}$	
Transition probabilities			
Reduced risk of developing DM2 for the intensive variant of the lifestyle-intervention program (hazard ratio)	$\beta^{\text{intensive}}$	0.70	[24]
Reduced risk of developing DM2 for the basic variant of the lifestyle-intervention program (hazard ratio)	β^{basic}	0.85	Expert judgment
Increased risk of death with diabetes (hazard ratio)	β^{diabetes}	2.13	[23]
Instantaneous rate of transiting from the *no diabetes* to the *diabetes* state in subpopulation kl	μ_{kl}^{12}	$\frac{-\log(1-I_{kl}(7))}{7}$	
One-year probability of transiting from the *no diabetes* to the *diabetes* state in subpopulation kl	$p_{kl}^{12}\left(t_{kl}^{i}\right)$	$1 - \exp\left(-\beta^{t_{kl}^{i}}\mu_{kl}\right)$	
One-year probability of transiting from the *no diabetes* to the *death* state (sex and age dependent)	$p^{13}(\text{sex}, \text{age})$	various	National life tables
One-year probability of transiting from the *diabetes* to the *death* state (sex and age dependent)	$p^{23}(\text{sex}, \text{age})$	$1 - \exp\left(\log(1 - p^{13}(\text{sex}, \text{age}))\beta^{\text{diabetes}}\right)$	
Costs and utilities attached to the Markov model's health states			
Cost attached to the *no diabetes* state		0	
Utility attached to the *no diabetes* state		0.84	[25]
Cost attached to the *diabetes* state		1805	[23]
Utility attached to the *diabetes* state		0.65	[23]
Costs of the lifestyle-intervention program			
Cost of the intensive variant in the first year		800	[24]
Cost of the intensive variant in the second and third years		520	[24]
Cost of the intensive variant in all subsequent years		0	
Cost of the basic variant in the first year		320	Expert judgment
Cost of the basic variant in the second and third year		160	Expert judgment
Cost of the basic variant in all subsequent years		0	
Patient characteristics			
Mean age in subpopulations 12 and 21		60	Expert judgment
Mean age in subpopulations 13 and 31		63	Expert judgment
Mean age in subpopulations 23 and 32		67	Expert judgment
Fraction of female subjects in subpopulation kl		0.543	[22]

developing DM2. Screening variant s^1 is based on existing clinical risk factors that have previously been shown to have a strong predictive power for the risk of developing DM2. Screening variant s^2 comprises the same risk factors as well as a hypothetical new biomarker for predicting the onset of DM2, such as a genetic marker related to metabolic programming by perinatal nutrition or a blood-based marker related to lipotoxicity and its metabolic consequences. As a result of the initial screening, individuals are classified into three risk categories. Individuals who are considered to be at high risk are offered an intensive, 3-year lifestyle-intervention program consisting of both a dietary part and a physical activity part. Individuals who are considered to be at intermediate risk are offered a more basic variant consisting of a dietary component only. No intervention is offered to individuals who are considered to be at low risk.

5.2. *Structure of the disease progression model*

To estimate the expected lifetime cost and QALY consequences of applying t^i_{kl} to subpopulation kl, we constructed a discrete-time Markov model with three health states (see Figure 2 for a schematic representation): *no diabetes*, *diabetes*, and *death*. We assumed the transition probabilities $p^{13}(\text{sex}, \text{age})$ and $p^{23}(\text{sex}, \text{age})$ to depend on sex and age, whereas we assumed the probability $p^{12}_{kl}\left(t^i_{kl}\right)$ of making a transition from the *no diabetes* to the *diabetes* state to depend on the subpopulation kl and on the treatment t^i_{kl} that is assigned to these individuals under screening variant s^i. The applied cycle length was 1 year.

5.3. *Parameter estimation*

Table II shows a summary of all parameters of interest, their initial values, and the sources used to obtain these values. As the novel biomarker included in s^2 had not yet been evaluated in a prospective cohort study, we had to rely on surrogate data to obtain initial values of some of these parameters. How this was carried out exactly is briefly described in the subsections in the following text.

5.3.1. Specification of the fractions f_{kl}. Salomaa *et al.* [22] evaluated whether a combined score of four novel biomarkers (adiponectin, apolipoprotein B, C-reactive protein, and ferritin) could improve the prediction of clinically incident diabetes over and above 11 classical risk factors, including blood glucose. For the purpose of this case study, we assumed that the performance of this biomarker score could serve as a proxy for the performance of the novel biomarker included under s^2. This allowed us to derive initial values of the fractions f_{kl} from the reclassification table that the authors produced for performing their NRI calculations (Table III) by setting $\gamma^1 = \gamma^2 = (0.03, 0.15)'$ and $f_{kl} = n_{kl}/n$.

5.3.2. Specification of the transition probabilities of the Markov model. We obtained the sex-dependent and age-dependent transition probabilities from the *no diabetes* to the *death* state from national life tables. We derived the transition probabilities from the *diabetes* to the *death* state from the transition probabilities from the *no diabetes* to the *death* state by first converting them into instantaneous death rates and then multiplying these death rates by a relative risk increase (hazard ratio) of $\beta^{\text{diabetes}} = 2.13$, which was obtained from a previously conducted economic evaluation [23]. To estimate $p^{12}_{kl}\left(t^i_{kl}\right)$, we assumed that the cumulative incidence functions $I_{kl}(\tau)$ were exponentially distributed, which allowed us to express the underlying instantaneous transition rates μ_{kl} as [16]

$$\mu_{kl} = \frac{-\log(1 - I_{kl}(\tau))}{\tau}. \tag{6}$$

Table III. Observed number of diabetes cases $\left(n^{\text{event}}_{kl}\right)$/total number of subjects (n_{kl}) after 7 years of follow-up in the HEALTH 2000 cohort [22].

		Predicted risk with classical risk factors plus biomarker score		
		Low risk	Intermediate risk	High risk
Predicted risk with classical risk factors	Low risk	29/3029	9/141	–
	Intermediate risk	8/337	89/1228	15/68
	High risk	0/1	5/27	33/146

We can then express the corresponding 1-year transition probabilities $p_{kl}^{12}\left(t_{kl}^{i}\right)$ as

$$p_{kl}^{12}\left(t_{kl}^{i}\right) = 1 - \exp\left(-\beta^{t_{kl}^{i}}\mu_{kl}\right), \tag{7}$$

where $\beta^{t_{kl}^{i}}$ denotes the hazard ratio comparing individuals receiving the preventive intervention t_{kl}^{i} to individuals not receiving a preventive intervention. For this case study, we set $\beta^{\text{intensive}} = 0.7$, which corresponds to the reduction in DM2 risk observed in the SLIM study [24]. As lifestyle interventions consisting of a dietary component only are less effective than lifestyle interventions consisting of both a dietary and a physical component, we assumed the value of $\beta^{\text{intermediate}}$ to be slightly higher and set equal to 0.85. We subsequently applied Equations (6) and (7) to transform the 7-year diabetes incidences $I_{kl}(7) = n_{kl}^{\text{event}}/n_{kl}$ as observed in the HEALTH 2000 cohort (Table III) into initial values of the 1-year transition probabilities $p_{kl}^{12}\left(t_{kl}^{i}\right)$.

5.3.3. Cost and utility estimates. We obtained the cost and utility estimates attached to the Markov model's transient health states from previously conducted economic evaluations. We derived the costs associated with the intensive variant of the lifestyle-intervention program from the SLIM study and set equal to EUR 800 for the first year and EUR 520 for the second and third years. For the basic variant of the lifestyle-intervention program, we set these costs equal to EUR 320 and EUR 160, respectively.

5.3.4. Specification of the patient characteristics. As we did not have access to the original data used to construct Table III, we had to rely on expert judgment to obtain initial values of the mean age and male/female ratio in the different subpopulations kl. Age is generally considered to be a strong predictor for the development of DM2, and this was taken into account when specifying the base-case values of the mean age in each of the subpopulations kl. In particular, we set the mean age in individuals who were classified as low risk under one of the screening variants and high risk under the other equal to 63, which corresponds to the third quartile of the age distribution observed in the HEALTH 2000 cohort. We subsequently assumed the mean age in individuals who are classified as low risk under one screening variant and intermediate risk under the other to be slightly lower and set equal to 60, whereas we assumed the mean age in individuals who are classified as intermediate risk under one screening variant and high risk under the other to be slightly higher and set equal to 67. Sex is usually not associated with the development of DM2. We therefore assumed the fraction of female subjects in each of the subpopulations kl to be equal to 0.543, which corresponds to the fraction of female subjects observed in the HEALTH 2000 cohort.

5.4. Results of the initial economic evaluation

For the base-case values of the model parameters, we found the commercial headroom available to the biomarker-based screening variant to be equal to EUR 75. To determine the robustness of this result with respect to small changes in the parameter values, we performed a one-way sensitivity analysis (Figure 3). The amount of headroom available was most sensitive to changes in the effect of the two lifestyle interventions and to changes in the mean age in each of the subpopulations kl. Changing the 7-year disease incidences $I_{kl}(7)$ had a less profound impact on $\Delta_{\text{scr}}^{\max}$. We also varied the male/female ratios in each of the subpopulations kl as well as the different cost components of the two lifestyle interventions, but these changes only had a marginal impact on the amount of headroom available (results not shown).

5.5. Implications

On the basis of the results of our initial economic evaluation, we can conclude that if the cost of measuring the novel biomarker is expected to be relatively low, there seems still sufficient room for improving the predictive performance of the existing risk classification models to warrant further research on novel biomarkers that are independently associated with the onset of DM2. On the other hand, if the unit cost of measuring the novel biomarker is likely to exceed EUR 100, it may be more fruitful to focus the research effort on identifying prognostic biomarkers that have a strong correlation with one of the established risk factors but are less expensive and/or invasive to measure. However, whether such a biomarker would actually be suitable as a substitute for an established risk factor not only depends on the effect size of this new biomarker compared with the established risk factor but also on its whole correlation

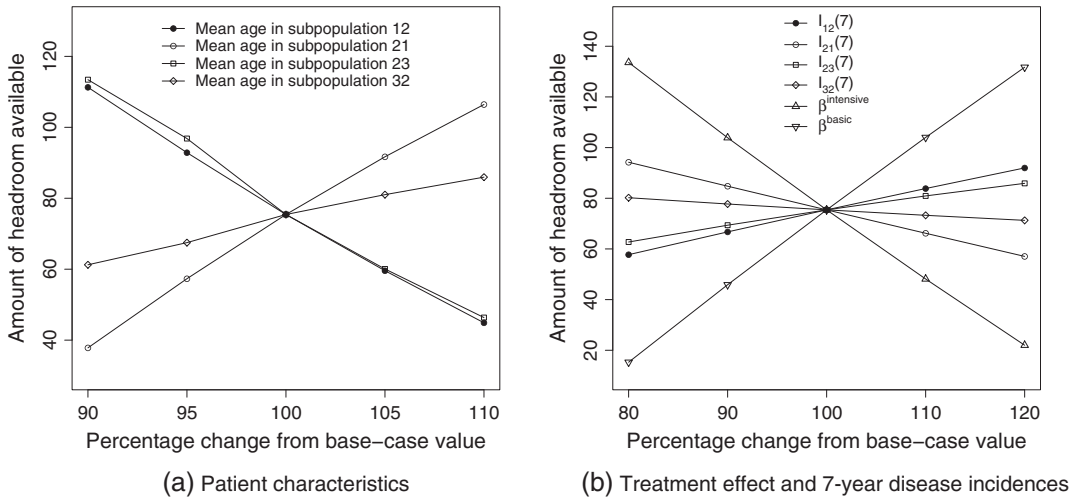

Figure 3. Results of the one-way sensitivity analysis.

structure with all the other variables included in the risk stratification model. This should be considered as well when making a go/no-go decision on the search for such a biomarker.

When performing our analysis, we implicitly assumed that the hypothetical new biomarker will have similar predictive capabilities as the biomarker risk score considered by Salomaa *et al.* [22]. It should therefore be noted that if such a biomarker is expected to have better (worse) capabilities in reclassifying subjects at risk of developing DM2, the cost of measuring the biomarker may be higher (should be lower) than the suggested upper bound of EUR 100. However, care should be taken not to raise the amount of headroom available too easily as the results of previous studies suggest that the initial expectations of a new biomarker have often been too optimistic, with disappointments in later phases of analyses [1].

6. Discussion

Moving from a screening variant based on traditional risk factors to a screening variant based on traditional risk factors plus a novel biomarker results in a reclassification of some of the individuals. To determine the maximum increase in screening cost for which this reclassification is still likely to be cost effective, we first restructured the decision problem in such a way that part of the parameters of interest could be estimated through disease progression modeling. We then described how these models could be combined with estimated values of the degree of reclassification to obtain initial estimates of the amount of commercial headroom available. We illustrated our method in a case study related to the prevention of DM2, where we used a Markov model with three health states to perform an initial economic evaluation of a potential new biomarker technology by using published data on the NRI of related but already clinically validated biomarkers.

Pietzsch and Paté-Cornell [2] have previously suggested a general method for the early HTA of new medical devices. Their approach requires an analyst to represent the dependency between the decision variable and the outcome of interest through a sequence of primary and intermediate effect variables, thereby allowing the analyst to obtain concrete realizations of the outcome variable by sampling from a series of conditional probability distributions. Our method is similar in the sense that we also determine the effect of the decision variable (biomarker-based screening or screening without using the biomarker) on the outcome of interest (the amount of headroom available) by sampling from a series of conditional probability distributions. However, instead of requiring the analyst to provide exact functional forms for each of these probability distributions, we have restructured the decision problem in such a way that some of these distributions can be approximated through disease progression modeling. This makes our method easier to apply in situations where there are no clear functional relationships between the variables of interest, such as in our case study related to the prevention of DM2.

A limitation of performing early HTA is that there is generally only a limited amount of clinical data available with which to populate the decision models. This implies that to be able to compute the amount of headroom available, it is sometimes required to make strong assumptions on some of the

unknown parameters, and in this respect, our method is no exception. In our case study, this became especially apparent when specifying the values of the cutoff points used to differentiate between low-risk, intermediate-risk, and high-risk individuals, the changes in the distribution of individuals across these three risk categories as a result of moving from the traditional risk classification model to the biomarker-based risk classification model, and the patient characteristics in each of the six subpopulations. The specification of the cutoff points should ideally be based on the ratio of the costs associated with a false positive and the benefits foregone because of a false negative [14], and the NMB framework provides a means of formally quantifying this trade-off. However, even with such a formal framework in place, it remains difficult to determine the values of the cutoff points in an 'objective' way: (i) analysts are still required to make a value judgment about the willingness to pay per unit of health gain, and (ii) the effectiveness of the administered treatments is likely to depend on the selected cutoff points, but the clinical data required to estimate this dependency may not be available.

Although the use of our method provides insight into the amount of headroom available to a novel biomarker, it does not directly provide an answer to the question of whether a medical technology firm should proceed with developing a technology that can measure this biomarker in actual clinical settings. To address this latter aspect, the results of the initial economic evaluation must first be translated into an estimate of the technology's maximum sales price by applying the principle of value-base pricing [26]. This estimate can then be fed into an appropriate product investment evaluation method, such as the one proposed by Girling *et al.* [27], to determine whether the expected post-market cash flows are sufficiently large to warrant further investments to transform the current concept into a fully developed end product. When performing such a return-on-investment analysis for a specific biomarker technology, one should be aware that the technology can potentially be used for multiple purposes, for example, not only as a screening test for selecting individuals eligible for a subsequent preventive intervention but also as a test for monitoring treatment response once the disease has been clinically established. All these potential uses of a new biomarker technology should ideally be taken into account when determining the technology's maximum sales price and estimating the subsequent expected post-market cash flows. However, performing such a comprehensive return-on-investment analysis is not straightforward, and future research on this problem area seems desirable.

To conclude, we presented a method for the early HTA of novel biomarker measurement in primary prevention programs and applied this method in a case study related to the prevention of DM2. Although we have focused on the use of the biomarker as a screening test for identifying individuals at risk of developing a chronic disease, our approach of first identifying the parameters of interest and then restructuring the decision problem in such a way that part of these parameters can be estimated through disease progression modeling seems more generally applicable. Future research effort may therefore be directed at exploring whether it is possible to quantify the clinical value of other potential applications of a new biomarker technology, such as a diagnostic test or a disease monitoring test, in a similar way.

Acknowledgements

This research was conducted within the framework of CTMM, the Center for Translational Molecular Medicine (www.ctmm.nl), projects PREDICCt (grant 01C-104), TRIUMPH (grant 01C-103), and PCCM (grant 03O-203), and supported by the Dutch Heart Foundation, Dutch Diabetes Research Foundation, and Dutch Kidney Foundation.

References

1. Hlatky M, Greenland P, Arnett D, Ballantyne C, Criqui M, Elkind M, Go A, Harrell F Jr, Hong Y, Howard B, Howard V, Hsue P, Kramer C, McConnell J, Normand S, O'Donnell C, Smith S Jr, Wilson P. Criteria for evaluation of novel markers of cardiovascular risk: a scientific statement from the american heart association. *Circulation* 2009; **119**(17):2408–2416. DOI: 10.1161/CIRCULATIONAHA.109.192278.
2. Pietzsch J, Paté-Cornell M. Early technology assessment of new medical devices. *International Journal of Technology Assessment in Health Care* 2008; **24**(1):36–44. DOI: 10.1017/S0266462307080051.
3. Chambless L, Cummiskey C, Cui G. Several methods to assess improvement in risk prediction models: extension to survival analysis. *Statistics in Medicine* 2011; **30**(1):22–38. DOI: 10.1002/sim.4026.
4. Chambless L, Diao G. Estimation of time-dependent area under the ROC curve for long-term risk prediction. *Statistics in Medicine* 2006; **25**(20):3474–3486. DOI: 10.1002/sim.2299.
5. Cook N, Ridker P. Advances in measuring the effect of individual predictors of cardiovascular risk: the role of reclassification measures. *Annals of Internal Medicine* 2009; **150**(11):795–802. DOI: 10.1059/0003-4819-150-11-200906020-00007.

6. Heagerty P, Zheng Y. Survival model predictive accuracy and ROC curves. *Biometrics* 2005; **61**(1):92–105. DOI: 10.1111/j.0006-341X.2005.030814.x.
7. Pencina M, D'Agostino R Sr, D'Agostino R Jr, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Statistics in Medicine* 2008; **27**(2):157–172. DOI: 10.1002/sim.2995.
8. Pencina M, D'Agostino R Sr, Steyerberg E. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Statistics in Medicine* 2011; **30**(1):11–21. DOI: 10.1002/sim.4085.
9. Vickers A, Elkin E. Decision curve analysis: a novel method for evaluating prediction models. *Medical Decision Making* 2006; **26**(6):565–574. DOI: 10.1177/0272989X06295361.
10. Hemingway H, Henriksson M, Chen R, Damant J, Fitzpatrick N, Abrams K, Hingorani A, Janzon M, Shipley M, Feder G, Keogh B, Stenestrand U, McAllister K, Kaski J, Timmis A, Palmer S, Sculpher M. The effectiveness and cost-effectiveness of biomarkers for the prioritisation of patients awaiting coronary revascularisation: a systematic review and decision model. *Health Technology Assessment* 2010; **14**(9):1–178. DOI: 10.3310/hta14090.
11. Henriksson M, Palmer S, Chen R, Damant J, Fitzpatrick N, Abrams K, Hingorani A, Stenestrand U, Janzon M, Feder G, Keogh B, Shipley M, Kaski J, Timmis A, Sculpher M, Hemingway H. Assessing the cost effectiveness of using prognostic biomarkers with decision models: case study in prioritising patients waiting for coronary artery surgery. *British Medical Journal* 2010; **340**:b5606. DOI: 10.1136/bmj.b5606.
12. IJzerman M, Steuten L. Early assessment of medical technologies to inform product development and market access: a review of methods and applications. *Applied Health Economics and Health Policy* 2011; **9**(5):331–347. DOI: 10.2165/11593380-000000000-00000.
13. Cosh E, Girling A, Lilford R, McAteer H, Young T. Investing in new medical technologies: a decision framework. *Journal of Commercial Biotechnology* 2007; **13**(4):263–271. DOI: 10.1057/palgrave.jcb.3050062.
14. Steyerberg E, Vickers A, Cook N, Gerds T, Gonen M, Obuchowski N, Pencina M, Kattan M. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010; **21**(1):128–138. DOI: 10.1097/EDE.0b013e3181c30fb2.
15. Drummond M, Sculpher M, Torrance G, O'brien B, Stoddart G. *Methods for the Economic Evaluation of Health Care Programmes*. Oxford University Press: New York, 2005.
16. Briggs AH, Claxton K, Sculpher M. *Decision Modelling for Health Economic Evaluation*. Oxford University Press: New York, 2006.
17. Groot Koerkamp B, Stijnen T, Weinstein M, Hunink M. The combined analysis of uncertainty and patient heterogeneity in medical decision models. *Medical Decision Making* 2011; **31**(4):650–661. DOI: 10.1177/0272989X10381282.
18. Garthwaite P, Kadane J, O'Hagan A. Statistical methods for eliciting probability distributions. *Journal of the American Statistical Association* 2005; **100**(470):680–701. DOI: 10.1198/016214505000000105.
19. Vallejo-Torres L, Steuten L, Parkinson B, Girling A, Buxton M. Integrating health economics into the product development cycle: a case study of absorbable pins for treating hallux valgus. *Medical Decision Making* 2011; **31**(4):596–610. DOI: 10.1177/0272989X10388041.
20. Briggs A, Ades A, Price M. Probabilistic sensitivity analysis for decision trees with multiple branches: use of the Dirichlet distribution in a Bayesian framework. *Medical Decision Making* 2003; **23**(4):341–350. DOI: 10.1177/0272989X03255922.
21. Gillies C, Abrams K, Lambert P, Cooper N, Sutton A, Hsu R, Khunti K. Pharmacological and lifestyle interventions to prevent or delay type 2 diabetes in people with impaired glucose tolerance: systematic review and meta-analysis. *British Medical Journal* 2007; **334**(7588):299. DOI: 10.1136/bmj.39063.689375.55.
22. Salomaa V, Havulinna A, Saarela O, Zeller T, Jousilahti P, Jula A, Muenzel T, Aromaa A, Evans A, Kuulasmaa K, Blankenberg S. Thirty-one novel biomarkers as predictors for clinically incident diabetes. *PLoS ONE* 2010; **5**(4):e10100. DOI: 10.1371/journal.pone.0010100.
23. Gillies C, Lambert P, Abrams K, Sutton A, Cooper N, Hsu R, Davies M, Khunti K. Different strategies for screening and prevention of type 2 diabetes in adults: cost effectiveness analysis. *British Medical Journal* 2008; **336**(7654):1180–1185. DOI: 10.1136/bmj.39545.585289.25.
24. Roumen C, Corpeleijn E, Feskens E, Mensink M, Saris W, Blaak E. Impact of 3-year lifestyle intervention on postprandial glucose metabolism: the SLIM study. *Diabetic Medicine* 2008; **25**(5):597–605. DOI: 10.1111/j.1464-5491.2008.02417.x.
25. Sullivan P, Ghushchyan V. Preference-based EQ-5D index scores for chronic conditions in the United States. *Medical Decision Making* 2006; **26**(4):410–420. DOI: 10.1177/0272989X06290495.
26. Claxton K, Briggs A, Buxton M, Culyer A, McCabe C, Walker S, Sculpher M. Value based pricing for NHS drugs: an opportunity not to be missed? *British Medical Journal* 2008; **336**(7638):251–254. DOI: 10.1136/bmj.39434.500185.25.
27. Girling A, Young T, Brown C, Lilford R. Early-stage valuation of medical devices: the role of developmental uncertainty. *Value in Health* 2010; **13**(5):585–591. DOI: 10.1111/j.1524-4733.2010.00726.x.